

Attentive and pre-attentive aspects of figural processing

Lawrence G. Appelbaum

Center for Cognitive Neuroscience,
Duke University, Durham, NC, USA



Anthony M. Norcia

Smith-Kettlewell Eye Research Institute,
San Francisco, CA, USA



Here we use the steady-state visual evoked potential (SSVEP) to study attentive versus non-attentive processing of simple texture-defined shapes. By “tagging” the figure and background regions with different temporal frequencies, the method isolates response components associated with the figure region, the background region, and with non-linear spatio-temporal interactions between regions. Each of these response classes has a distinct scalp topography that is preserved under differing attentional task demands. In one task, attention was directed to discrimination of shape changes in the figure region. In the other task, a difficult letter discrimination was used to divert attentive processing resources away from the texture-defined form. Larger task-dependent effects were observed for figure responses and for the figure/background interaction than for the background responses. The figure region responses were delayed in occipital areas in the shape versus letter task conditions, while the region interactions were enhanced, especially in frontal areas. While a basic differentiation of figure from background processing occurred independent of task, attentive processing of elementary shapes recruited later occipital activity for figure processing and sustained non-linear figure/background interaction in frontal areas. Collectively, these results indicate that basic aspects of scene segmentation proceed pre-attentively, but that directed attention to the object shape engages a widely distributed network of brain areas including frontal and occipital regions.

Keywords: attention, figure–ground segmentation, EEG, non-linear spatio-temporal interactions

Citation: Appelbaum, L. G., & Norcia, A. M. (2009). Attentive and pre-attentive aspects of figural processing. *Journal of Vision*, 9(11):18, 1–12, <http://journalofvision.org/9/11/18/>, doi:10.1167/9.11.18.

Introduction

Visual scene segmentation is the cognitive ability to group stimulus features within distinct objects and surfaces and to segregate these features when they arise from different objects or surfaces in a manner that promotes a biologically useful representation of the visual scene. At least some aspects of this process can proceed automatically without the participant’s awareness (Kastner, De Weerd, & Ungerleider, 2000; Schira, Fahle, Donner, Kraft, & Brandt, 2004; Scholte, Witteveen, Spekrijse, & Lamme, 2006; Schubo, Meinecke, & Schroger, 2001). However, it has also been reported that conscious processing of scene attributes involves additional, longer latency processing after stimulus presentation (Boehler, Schoenfeld, Heinze, & Hopf, 2008; Caputo & Casco, 1999; Casco, Grieco, Campana, Corvino, & Caputo, 2005; Fahrenfort, Scholte, & Lamme, 2008; Heinrich, Andres, & Bach, 2007; Scholte et al., 2006; Schubo et al., 2001), possibly occurring in higher order visual areas (Kastner et al., 2000; Schira et al., 2004; Scholte et al., 2006).

Figures or real-world objects only exist in the context of their relationship to a background or supporting surfaces, such as the ground plane. Conscious processing of object attributes could therefore act selectively to enhance processing of the figure itself, or it could involve

suppression of the background region, or both. Beyond this, conscious processing may also affect spatio-temporal neural interactions between the two regions that are organized either at the level of feature discontinuities across regions, such as end-stopping, or at the object level. In this study, we sought evidence for each of these possible response modifications by using an EEG temporal-tagging method in which figure and background regions are modulated at distinct temporal frequencies. So-called, steady-state visually evoked potentials (SSVEP) arising from each of these regions can be recovered via spectral analysis, even when both regions are present. Non-linear spatio-temporal interactions between regions can also be detected by monitoring responses at frequencies equal to low-order sums and differences of the region tag frequencies.

Using the temporal tagging method, we have previously shown that the figure and background regions of simple second-order texture stimuli activate distinct processing networks, with the figure region preferentially engaging lateral occipital areas associated with object processing (Kourtzi & Kanwisher, 2000; Malach et al., 1995), while the background preferentially engaged more dorso-medial areas (Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006). Within this paradigm, we were also able to isolate response components that reflect non-linear spatio-temporal interactions between the figure and background regions (Appelbaum, Wade, Pettet, Vildavski, & Norcia, 2008).

These interaction components were shown to receive strong contributions from texture elements at the border of the figure and background regions, carrying spatially precise, time-locked information about the continuity/discontinuity of textures in the two regions.

Frequency-tagged stimuli have also been used to study spatial attention (Chen, Seth, Gally, & Edelman, 2003; Ding, Sperling, & Srinivasan, 2006; Kim, Grabowecky, Paller, Muthu, & Suzuki, 2007; Morgan, Hansen, & Hillyard, 1996; Müller, Teder-Sälejärvi, & Hillyard, 1998) and feature-based attention (Müller et al., 2006; Pei, Pettet, & Norcia, 2002). In these studies, the response to the attended location or feature was larger than when that location or feature was not task relevant, suggesting that attention serves to increase the sensory gain for the attended stimulus.

By using simple texture-defined forms that contain only a limited number of cues on which figure–ground segmentation can occur and by isolating qualitatively meaningful response components, we can precisely control and monitor the ways in which figure and background networks are modulated by specific task-dependent attentional demands. To assess the role of active attentive processing on the dynamics of figure–ground networks we recorded evoked responses, using whole-head EEG, while participants performed two different demanding behavioral tasks, one in which scrutiny of the region shapes was necessary for successful performance and one in which attention was directed to a stream of concurrently presented letters. We found that the timing (phase) of the figure region response was preferentially modulated by task compared to the background response, as was the magnitude of certain components of the figure–background interaction. Strikingly, sustained task-dependent non-linear figure–ground interaction was found over frontal cortex in addition to more transient interactions found over occipital cortex, demonstrating that attentive processing of object shapes involves a widely distributed network of occipital and frontal areas.

Methods

Participants

Twelve normal, neurologically intact adults (28–58 yrs), with normal or corrected-to-normal visual acuity, participated in this experiment. One subject was run twice to informally evaluate retest consistency. Though both data sets showed consistent results, only the first session was included in subsequent analyses. Informed consent was obtained prior to experimentation under a protocol that was approved by the institutional review board of the California Pacific Medical Center and conformed to the tenets of the Declaration of Helsinki.

Visual stimuli and tasks

Scene segmentation involves, minimally, the segmentation of a single object on a background or the separation of two surfaces. In natural scenes, there are usually many cues on which the segmentation can be based. Here we restricted the available cues to relative orientation and timing differences. A schematic illustration of the stimuli is shown in Figure 1a. The orientation of the texture within the figure region was updated at 3 Hz and that of the background at 3.6 Hz. This updating procedure yields four stimulus states: two uniform ones and two segmented ones.

All stimuli were comprised of one-dimensional random-luminance bars with a minimum bar width of 6 arc min.

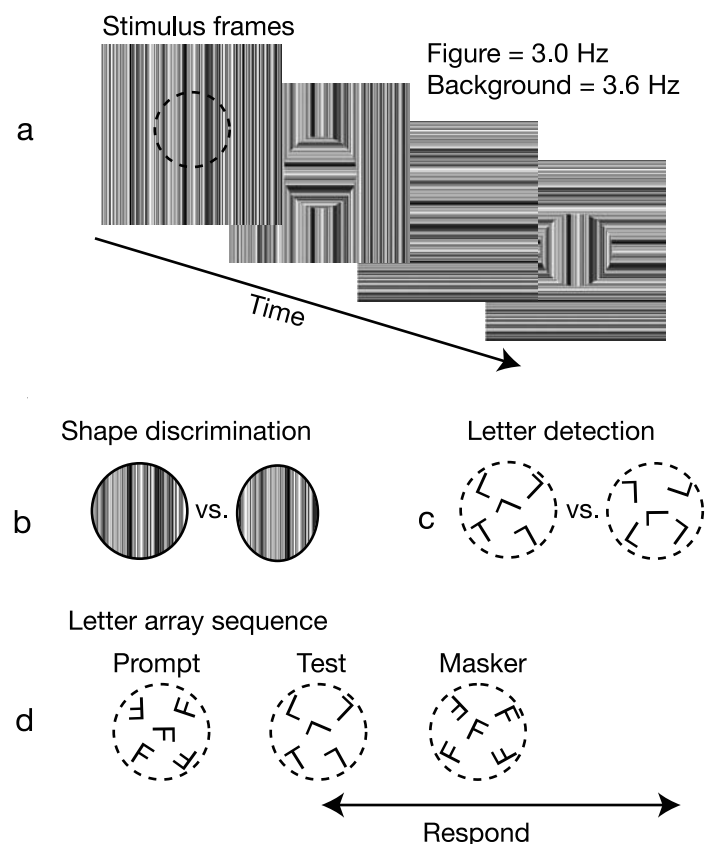


Figure 1. (a) Stimulus schematic illustrating the four frames of orientation defining the stimulus. The figure region alternated orientation between vertical and horizontal at 3.0 Hz, while the background alternated at 3.6 Hz. (b) Representative circular versus elliptical comparison of figure shape in the shape discrimination task. (c) Representative letter task comparison in which test arrays contain the target letter “T” (left) and target-absent arrays composed of all “L”s. (d) Order of stimulus arrays presented during the letter detection task. Each trial begins with a prompt array of all “F”s and ends with a masker array of all “F”s; test arrays presented between the prompt and masker stimuli are on a duration threshold based on subject performance. See [Supplementary material](#) for stimulus animation.

The maximum contrast between elements was 80%, based on the Michelson definition. The mean luminance was 56.3 cd/m^2 and the full display subtended $20^\circ \times 20^\circ$ of visual angle. Stimuli were designed such that the textures composing “figure” and “background” regions modulated at different temporal frequencies. The figure consisted of a circular 5° -diameter region, centered on fixation, which changed orientation by 90° at 3.0 Hz. The background texture, comprising the remainder of the texture, changed orientation by 90° at 3.6 Hz. Because the figure and background regions changed at different rates, the texture alternated between segmented and uniform states, where the figure existed only when the two regions were of dissimilar orientation, and the segmentation disappeared when they were both in alignment. The figure region was thus cued by both a difference in orientation and by a difference in the frequencies of the tags imposed on the two regions.

To assess the influence of attention on the amplitude and timing of brain responses involved in figure versus background processing, participants were instructed to fixate on a fixation mark at the center of the display and perform one of two different tasks in separate experimental blocks. In the *Shape Discrimination* trials, the observers indicated a change in the shape of the figure with a button press (see Figure 1b). On 20% of the 1.67-second stimulus cycles, the aspect ratio (horizontal to vertical) of the figure became elliptical (<1). Responses were monitored and the figure aspect ratio was adjusted to maintain performance at approximately 80% correct detection. In the *Letter Detection* trials, the observers were instructed to attend to a stream of simultaneously presented letters and to detect a probe letter “T” among distracters “L” (see Figures 1c and 1d). Letter arrays containing “L”s and “T”s appeared superimposed upon the figure region of the stimulus and were preceded and followed by masking arrays of only “F”s. On trials in which the target letter “T” appeared subjects pressed one button. On trials in which no target appeared, they pressed a second button. A staircase procedure was used to maintain a constant, high level of task difficulty by holding performance at or near the duration threshold for letter detection. The duration of presentation started at 1200 msec and was decreased by 10 msec with each correct response, and lengthened by 20 msec with each incorrect response according to an adaptive rule that converged on the 82% correct level of the psychometric function. Importantly, while the temporal tag of the figure region was periodic at 3.0 Hz, the letter presentation was broadband (non-periodic) and variable due to changes in the presentation duration used in the staircase procedure. Therefore, while a very small percentage of the stimulus modulation energy for the figure region and letter task were in phase, the vast majority of the energy was out of phase. Both tasks were present on the screen at the same time and only the instructions given to the observers differed between conditions. For each trial, correct

responses were defined by a button press occurring within 2 seconds of the target presentation, and condition order and button assignment were counterbalanced across subjects. Experimental runs lasted roughly 5 minutes and between four and six runs were collected on each participant.

EEG acquisition and statistical analysis

The electroencephalogram (EEG) was collected with a whole-head, 128-channel, Geodesic EEG system with HydroCell Sensor Nets (Electrical Geodesics, Eugene OR), amplified at a gain of 1,000, and recorded with a vertex physical reference. Signals were 0.1 Hz high-pass and 200 Hz low-pass filtered and digitized at 432 Hz with a precision of 4-bits per microvolt at the input. Artifact rejection and spectral analysis of EEG data were done off-line. Raw data were evaluated according to a sample-by-sample thresholding procedure to remove noisy sensors which were replaced by the average of the six nearest spatial neighbors. Once noisy sensors were substituted, the EEG was re-referenced to the common average of all the sensors. Additionally, EEG epochs that contained a large percentage of data samples exceeding threshold (~ 25 – $50 \mu\text{V}$) were excluded on a sensor-by-sensor basis. Time-locked averaged waveforms for each stimulus condition were computed over one stimulus cycle (1.67 seconds) for each subject and for the grand average across subjects. Individual subject and group time averages were converted to complex-valued amplitude spectra at a frequency resolution of 0.6 Hz via a discrete Fourier transform (DFT). The resulting amplitude spectra of the SSVEP were then evaluated at a set of frequencies uniquely attributable to the input stimulus frequency tags up to the 18th and 15th harmonic for the figure and background tags, respectively.

Because periodic visual stimulation leads to a periodic visual response occurring at frequencies equal to exact integer multiples of the stimulation frequency (Regan, 1989), individual spectral harmonics can be uniquely attributed to each region of the visual input. Several steps were taken to ensure that the evoked responses were exactly centered within single bins of the spectrum analysis. First, the visual display updates were performed in synchrony with the 72-Hz vertical retrace of the display. Secondly, a clock signal for the A/D conversion (432 Hz; six samples per video frame) was derived by down-counting the horizontal synch of the video monitor by exactly 101 video lines per A/D sample using a hardware device. Finally, the analysis epochs contained an exact integer number of stimulus cycles at both the 3.0- and 3.6-Hz stimulus rates (24 and 20 video frames, respectively). In our experiment, all frequencies of interest were exact integer submultiples of the 1.67-Hz common stimulus cycle (0.6 Hz or 120 video frames). A rectangular data window was used for the DFT. The steps outlined above center the response frequencies on the

frequency comb of the transform. Because each response frequency is precisely bin-centered, there is no leakage of energy from responses at one frequency into the other frequency bins. This is because the sinc function data window around each spectrum line has a zero crossing at each of the other spectrum bin locations. The narrow-band nature of both region-related and region interaction SSVEP response components was verified by performing the DFT at 0.06-Hz resolution for a representative sample of channels and participants (see examples in [Supplementary Figures 1 and 2](#)).

The time course of activity elicited by the figure and background regions was reconstructed by selectively back-transforming (via an inverse discrete Fourier transform) either the figure ($nf1$) or background ($mf2$) harmonics up to 54 Hz. Components that could be attributed to both figure and background were excluded, (e.g., $6f1 = 5f2 = 18$ Hz). This inverse transformation acts as a filter that enables examination of the temporal waveform of the SSVEP, while still allowing us to separate the signals generated by the two stimulus frequency tags (see [Figures 5 and 6](#)). Similarly, the time course of activity evoked by the figure and background region interactions was computed by selectively back-transforming all low-order sum and difference terms up to 54 Hz. This transform results in a complex waveform that reflects the temporal interaction between the two inputs and is only interpretable on the full 1.67-second stimulus cycle.

The statistical significance of individual participant responses at a given frequency in [Figure 2](#) and the significance of differences between conditions in group data in [Figure 4](#) were evaluated using the T_{circ}^2 statistic (Victor & Mast, 1991). This statistic utilizes both amplitude and phase consistency across trials (or subjects) to assess whether stimulus-locked activity is present that differs from noise and whether these responses differ between condition. For this purpose, the sine and cosine values are assumed to be independent and are pooled in quadrature to arrive at a standard error for the complex valued data. Standard error of the mean (SEM) measurement of this type is displayed as gray circles in [Figure 4](#).

Task-dependent group differences at the individual response frequencies were tested using permutation methods derived by Greenblatt and Pflieger (2004) and provided by the EMSE software package (Source Signal Imaging, San Diego, CA). Paired comparisons were first made on the complex-valued responses in order to test for amplitude and/or phase differences between the two experimental conditions under the null hypothesis that no differences exist and that the data are interchangeable. Here shape- and letter-task condition assignments were randomized across all subjects to compute a permutation sample distribution with 10,000 total samples of the vector difference between conditions.

$$|A - B| = \sqrt{\text{Re}(A - B)^2 + \text{Im}(A - B)^2} \quad (1)$$

where A and B are the single harmonic values for the two task conditions and Re and Im are the real and imaginary parts. Here the difference depends on both the amplitudes of A and B, and the angle between them and the test is against zero difference. The observed data differences were then compared to the permutation sample on a channel-by-channel basis.

To control Type II errors (false alarms), only frequency bins where more than 5% of the channels ($n = 6$) exceeded a $p < 0.05$ significance criterion were considered to have an omnibus significant difference between conditions. The histogram of significant differences reported in [Figure 3d](#) reflects this correction.

Results

Texture-defined forms and behavioral performance

Participants performed alternating runs in which they were asked to either detect subtle changes in the shape of the figure ([Figure 1b](#), shape discrimination) or to detect a target letter “T” in a concurrent and superimposed stream of letter arrays ([Figure 1c](#), letter detection). Both the shape and letter tasks were challenging and required focused attention to perform. In the shape discrimination task the mean shape change, defined as the horizontal-to-vertical aspect ratio, was 0.93. Subjects correctly identified 79.9% of these targets. In the letter task, the median value at which the target duration function converged (at 82% correct) was 362 msec.

Steady-state response spectra

Because the figure and background regions of our stimuli were tagged with different temporal frequencies, we are able to uniquely measure the responses generated by each region as well as the responses that correspond to their non-linear interactions. To illustrate the steady-state responses recorded in these tasks, [Figure 2](#) shows the amplitude spectrum at a single channel, for one observer in the shape discrimination condition.

As shown in [Figure 2](#), statistically significant steady-state responses were present at harmonics of the stimulus frequency tags and at frequencies equal to low-order sums and differences of the two tag frequencies. Statistical significance of individual records, such as this, was determined by the T_{circ}^2 (Victor & Mast, 1991). The responses evoked by the figure region are shown in blue (referred to as $nf1$ to indicate harmonic index n of stimulus frequency 1) while those evoked by the background are shown in green ($mf2$). Consistent with the

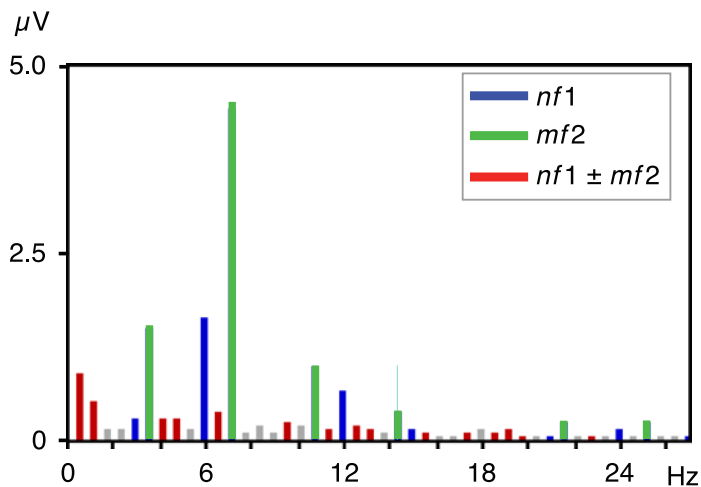


Figure 2. EEG amplitude spectra from the shape discrimination task for a single channel from one observer. Harmonics of the figure frequency tag are indicated in blue, background responses in green, and non-linear interactions in red.

presence of two temporal transients in each cycle (vertical and horizontal), the amplitude of the region responses are largest at the second harmonics of the stimulus tags. In addition, responses that reflect non-linear interactions between regions are present at frequencies equal to low-order sums and differences ($nf1 \pm mf2$) of the two tags and are highlighted in red. Those frequency components colored in gray did not meet the statistical criterion of $p > 0.05$ (T^2_{circ} test).

Using a variant of this tagging procedure, we previously found that the response topography and source distribution

of figure-related activity differed from that of background-related activity, and that this activity was independent of the texture cue or spatial layout used to elicit the response (Appelbaum et al., 2006). Additionally, we demonstrated that the $1f1 + 1f2$ non-linear interaction elicited by these displays was sensitive to the specific configuration of the figure-ground arrangement and was therefore an index of the border-related activity (Appelbaum et al., 2008).

The figure and background responses in the present data replicated the basic spatial disassociation between the region responses that was observed in our previous reports. As previously reported, activity elicited by the figure ($2f1$) produced a bilateral occipital distribution, while the background ($2f2$) produced a focal, midline posterior distribution. This spatial distinction was present irrespective of the attentional demands imposed by the observer's task and can be observed in the profile of waveform amplitudes in Figure 5.

Quantifying task-dependent modulations

Two types of analyses were carried out in order to evaluate amplitude and timing modulations separately for the figure, background, and figure-ground interaction responses that occur within texture-defined scenes. Analyses were first carried out on the complex-valued responses to arrive at an omnibus statistical summary over all electrodes and response harmonics. Next, the amplitude and phase of responses occurring at the individual harmonics were assessed separately in order to disassociate attentional effects on gain and timing of the evoked response for each region type. Finally, in order to

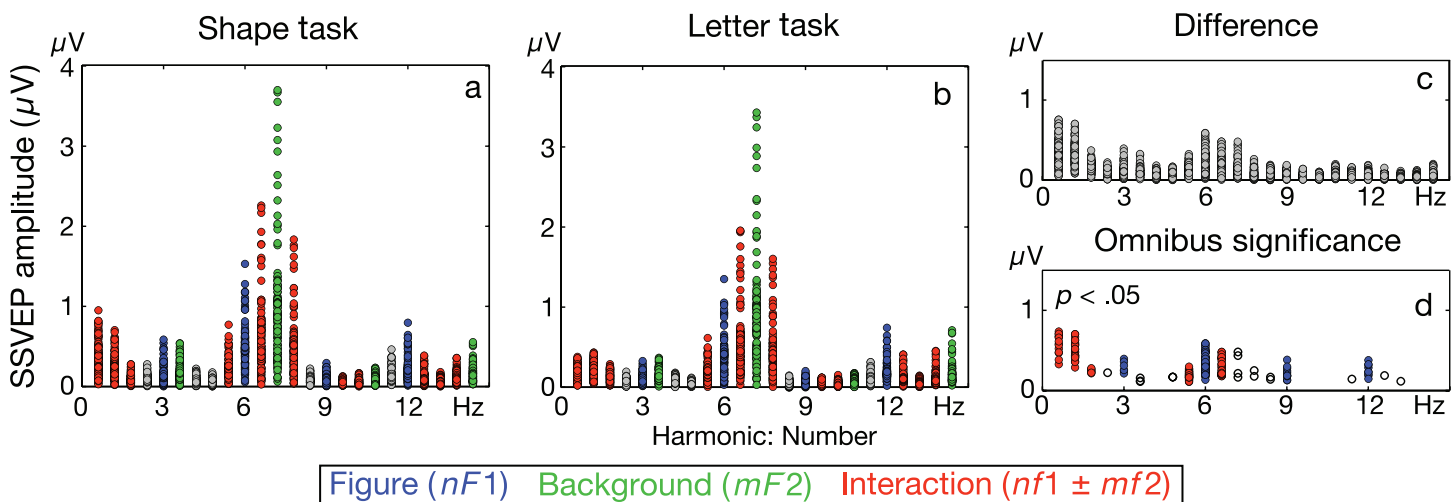


Figure 3. Amplitude spectra for the shape (a) and letter (b) tasks are shown as overlays of all 128 channels for the 12-observer grand average. The blue, green, and red colors are used to highlight the harmonic responses for the figure ($nf1$), background ($mf2$), and non-linear interaction terms ($nf1 \pm mf2$), respectively. The algebraic difference of all channels is shown in panel c. Channels that differ significantly between task conditions according to permutation tests ($p < 0.05$) are indicated in panel d. Those harmonics that reach significance according to an omnibus correction ($>5\%$ of channels) are colored. Only figure and interaction responses reach omnibus significance.

visualize the spatial distribution of these effects over the full 128-channel sensor array, reconstructed waveforms are presented (see [Figures 5 and 6](#)) that aggregate activity over all of the response components for each region type (figure = $1/f_1 + 2/f_1 + 3/f_1 + 4/f_1 \dots$) and for the region interaction.

Task-dependent modulations in the frequency domain

The effect of task was first evaluated on a frequency-by-frequency basis and with reference to whether a given frequency component was driven by the figure region, the background region, or was a result of non-linear spatio-temporal interaction. [Figures 3a and 3b](#) show the grand average amplitude spectra for the shape and letter task, respectively. Individual sensor data are plotted as separate circles in the spectra and colored according to their correspondence with the driving stimuli, using the same convention as presented in [Figure 2](#). Prominent responses are visible at several harmonics of the figure (blue) and background (green) region tag frequencies, and at their interaction frequencies (red), with the greatest amplitudes occurring at the second harmonics and the second-order sum frequency.

[Figure 3c](#) shows the algebraic difference of the response distribution for all channels, as filled gray circles. [Figure 3d](#) re-plots these data only for channels that reach permutation significance levels of $p < 0.05$. Those frequency components with greater than 5% of channels (>6) reaching significance are colored to indicate omnibus significance when corrected for type II errors across the whole map. Interestingly, only figure (blue) and interaction (red) response harmonics reach omnibus significance. These frequency components also show the largest proportional amplitude differences between task conditions.

Separating phase and amplitude effects on the SSVEP response

The SSVEP is a complex-valued quantity with both amplitude and phase information. Response amplitude is related to aggregate neuro-electric activity and therefore gives information about the response gain of activated neural populations that synchronize with the tagged stimuli. Phase is related to the relative delay of the responses, providing information about response timing.

The omnibus, whole map permutation statistics presented in [Figure 3](#) was derived from the complex-valued responses. The differences between conditions could thus be due to either phase or amplitude effects or to a combination of the two. As a preliminary analysis, we plotted the complex-valued amplitude and phase values of the SSVEP response for the 12-subject group average

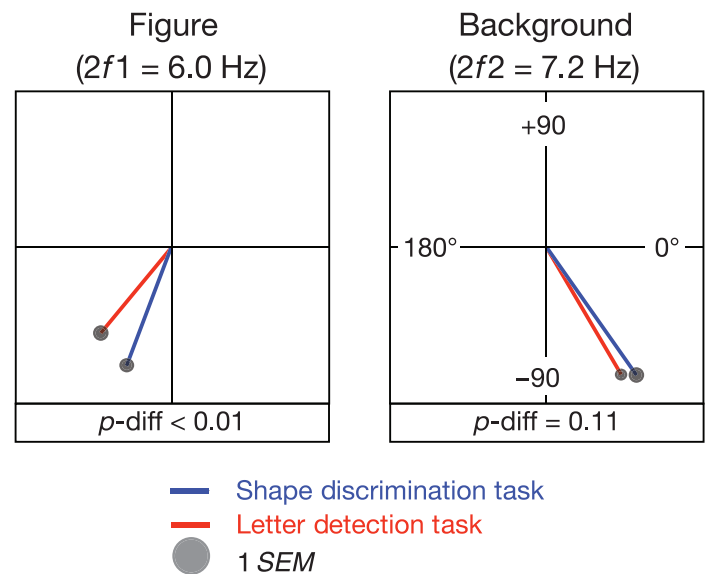


Figure 4. Fourier components of the 12-subject grand average SSVEP at frequencies $2f_1$ (left) and $2f_2$ (right). Single sensors, corresponding to the maximum amplitude, are plotted in the complex plane with the length of each vector representing the response amplitude and the direction representing phase. The radius of the gray circles at the end of each vector represents one standard error of the mean. Consistent phase delays (counterclockwise lags) are observed for the figure ($p < 0.01$), but not the background ($p = 0.11$) response when the task is to monitor the region shape.

occurring at the distribution maximum for each frequency component for the dominant figure-related ($2f_1$) and background-related responses ($2f_2$). In these plots, the vector length codes response amplitude and polar angle codes response phase for the shape discrimination task in blue, and the letter task in red. The phase convention in these plots places 0° delay (0 phase lag with respect to the stimulus) at 3 o'clock. Increased response lag runs in the counterclockwise direction and decreasing lag in the clockwise direction, relative to the time origin at 0° . Gray circles indicate one SEM and correspond to experimental errors of about 2 to 4 msec here. The error estimates were calculated by first computing the separate means and standard errors of the real and imaginary components over subjects and then pooling the errors in quadrature. The radius of the error circle corresponds to 1 SEM on this measure.

As is evident in [Figure 4](#), task demands reliably influence the figure region response amplitude and phase ($p\text{-diff} < 0.01$), as indicated by the T^2_{circ} test for mean vectors (Victor & Mast, 1991). The background region responses, on the other hand, do not differ in the two task conditions ($p\text{-diff} = 0.11$). For both regions, response amplitudes are about 8% larger during performance of the shape task, but for the figure, response phase is also

Harmonic	Frequency (Hz)	Amplitude ratio	Amplitude <i>p</i> values	Phase <i>p</i> values
$1f_2 - 1f_1$	0.6	2.44	<0.01*	0.40
$2f_2 - 2f_1$	1.2	1.43	0.17	0.86
$2f_1$	6	1.08	0.1	<0.01*
$1f_1 + 1f_2$	6.6	1.33	<0.01*	0.57
$2f_2$	7.2	1.08	0.08	0.07
$4f_1$	12	1.05	0.45	0.81
$2f_1 + -2f_2$	13.2	1.38	0.07	0.96
$4f_2$	14.4	1.02	0.71	0.29

Table 1. Summary of response modulation at second- and fourth-order harmonics and interactions.

delayed (is more counterclockwise) in the shape discrimination task relative to the letter task. The mean phase change due to attention for the figure region in these averages is equivalent to 8 msec while that for the background is 1.9 msec. This analysis again suggests that attentional demands preferentially affect the figure-region responses, leaving the background responses relatively unchanged.

To distinguish gain and timing contributions to the omnibus, whole map significance test of Figure 3, we evaluated the phase and amplitude effects separately for the two tasks at the second and fourth harmonics of the stimulus tags ($2f_1$, $2f_2$, $4f_1$, $4f_2$) and at the dominant interaction frequencies ($1f_2 - 1f_1$, $2f_2 - 2f_1$, $1f_1 + 1f_2$, $2f_2 - 2f_1$). Response gain ratios were computed as the maximum amplitude value recorded in the shape task divided by the amplitude at the same channel in the letter task. These values, shown in Table 1, are greater than 1 at all harmonics indicating that focused attention increases the gain of these responses but reach statistical significance only for a subset of the interaction components ($1f_2 - 1f_1$ and $1f_2 + 1f_1$). The amplitude ratios for these two components are 2.66 and 1.42, respectively. Because the total response to the display is a summation of the three classes of response (figure, background, and interaction), the overall response is larger during the shape than letter tasks. Conventional ERP analyses record the sum of these distinctive response classes and therefore are unable to isolate the amplitude effect as specific class of response as we do here.

Task-related phase differences, computed via two-tailed comparison of the phase difference between task conditions, reached significance for the second harmonic of the figure but was only marginally significant for the background (see right column of Table 1). The absolute latency of the phase modulation was nearly five times as large for $2f_1$ as $2f_2$ (-12.9° or 5.9 msec for $2f_1$; -3.2° or -1.2 msec for $2f_2$). Sizable, but non-significant phase differences are present for the low differences (-16.1° or -7.4 msec for $1f_2 - 1f_1$; 8.3° or 2.0 msec for $2f_2 - 2f_1$), and small, non-significant phase differences are present

for the sums (-1.5° or <-1 msec for $1f_1 + 1f_2$; 1.6° or <-1 msec for $2f_2 - 2f_1$).

Task-dependent modulations in the time domain

Each class of neural response in our tagging paradigm is comprised of several frequency components (e.g., the components analyzed individually in Figure 3 and Table 1). In order to visualize the influence of task demands on the full response from the figure, background, and interaction response classes, we applied an inverse Fourier transform to the relevant components separately within each response class (e.g., $nf_1 = \text{iFFT}[1f_1 + 2f_1 + 3f_1 + 4f_1 \dots]$). This accomplishes two things: first, it provides an aggregate measure of the total response to each region as well as the inter-region interaction; second, it provides an alternative, time-domain view of the data. The grand averages of these “filtered waveforms” are shown in Figure 5 separately for each response category. The waveforms are shown for all 128 channels in a topographic spatial arrangement corresponding to the sensor positions with selected channels, indicated by the boxes, enlarged to highlight task effects. The EEG trace is shown for one cycle of the figure (333 msec—left), one cycle of the background (277 msec—middle), and over one full cycle of the stimulus for the interactions (1.67 seconds = 5×333 msec = 6×277 msec—right). Waveforms for the shape discrimination task are shown in blue; waveforms for the letter task are shown in red.

The responses for both the figure and background regions are largest over occipital electrode sites and contain two large peaks per stimulus cycle, consistent with the observation that the second harmonic is the dominant response component for both regions (see Figure 2). In absolute terms, the background responses are much larger than the figure responses (see scale values to the left of each enlarged channel) which is not surprising given that they are generated by a stimulus area that is about six times larger than the figure regions. The phase difference between the two task conditions seen in the figure region frequency domain plots (Figure 4) is also evident as a shift of the zero crossings of the filtered figure waveform, and this effect is less apparent for the background response, consistent with the single component analyses of Figure 4 and Table 1. Recall that the magnitude of the phase effect was five times larger for the dominant component of the figure region response compared to the background response. For both the figure and background responses, amplitude effects are small in magnitude as they were in Table 1.

The response components derived from the interaction terms show a more complex pattern, with predominantly low-frequency activity occurring over frontal and fronto-lateral electrodes and higher frequency components

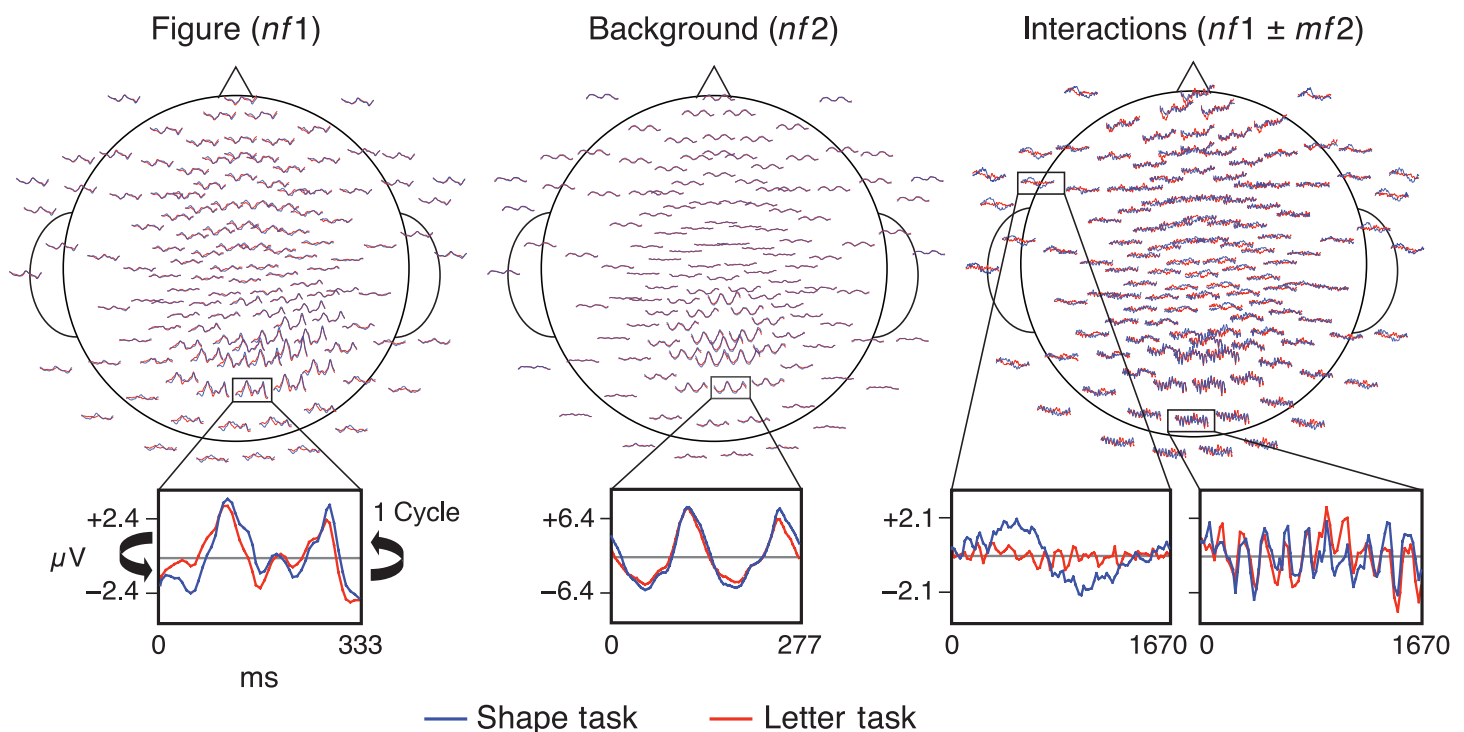


Figure 5. Grand average response waveform for one cycle of the figure ($nf1$), the background ($mf2$), and the non-linear interactions ($nf1 \pm mf2$). Waveforms are computed by selectively back transforming frequencies corresponding only to harmonics of the input tag or their low-order sums and differences. Waveforms for the shape discrimination task are shown in blue and the letter task in red. Figure and background waveforms show biphasic response with maximal activity over occipital sensors, while interactions contain low-frequency modulations over frontal and central sensors and higher frequency activity over posterior recording sites. See [Supplementary Figures 4a, 4b, and 4c](#) for enlarged images.

occurring over occipital sites. The omnibus test suggests that the $1f2 - 1f1$ and $2f1 - 2f2$ terms dominate the slow waveforms at frontal sites, and the $1f1 + 1f2$ component dominates over occipital areas.

Discussion

We have measured the responses to the figure and background regions of simple figure-ground displays using the frequency-tagged SSVEP and have found that the timing of the figure region is delayed by about 6 msec when figure shape is actively processed. The background region, by contrast, did not show a reliable response timing effect. The scalp topography of figure and background responses, and thus their underlying generator structure, was largely unaffected by attentional task, as we have reported previously (Appelbaum et al., 2006). Both figure and background networks show small enhancements of response amplitude when figure shape is actively processed, but these were not significant. Because the amplitude effects were non-significant, we do not find evidence for attention to the figure region resulting in a suppression

of the background response, which is one possible mechanism by which object regions could become more salient than background regions. It is possible that portions of the background response originating near the border between regions could have been depressed and that this effect may have been swamped by the responses from the remainder of the large surrounding backgrounds we used.

We observed significant task-related modulation of the non-linear interaction between the figure and background regions. This interaction is comprised of two sets of components with distinct scalp distributions (Figure 5); a high-frequency interaction that is visible over the posterior scalp at electrodes where the region responses are maximal and a low-frequency interaction that is largest at frontal and frontal-lateral electrodes. The magnitude of the low-frequency interaction is more highly modulated by task demands than is the high-frequency interaction or the region responses. The presence of distinct topographic distributions for the low- and high-frequency interactions means that they cannot be generated by the same source; rather, there appear to be two networks involved in processing the figure-ground interaction—one that is highly sustained and involves areas distant from the occipital areas and the other which is temporal bandpass (transient) that is centered in visual cortex. Current source

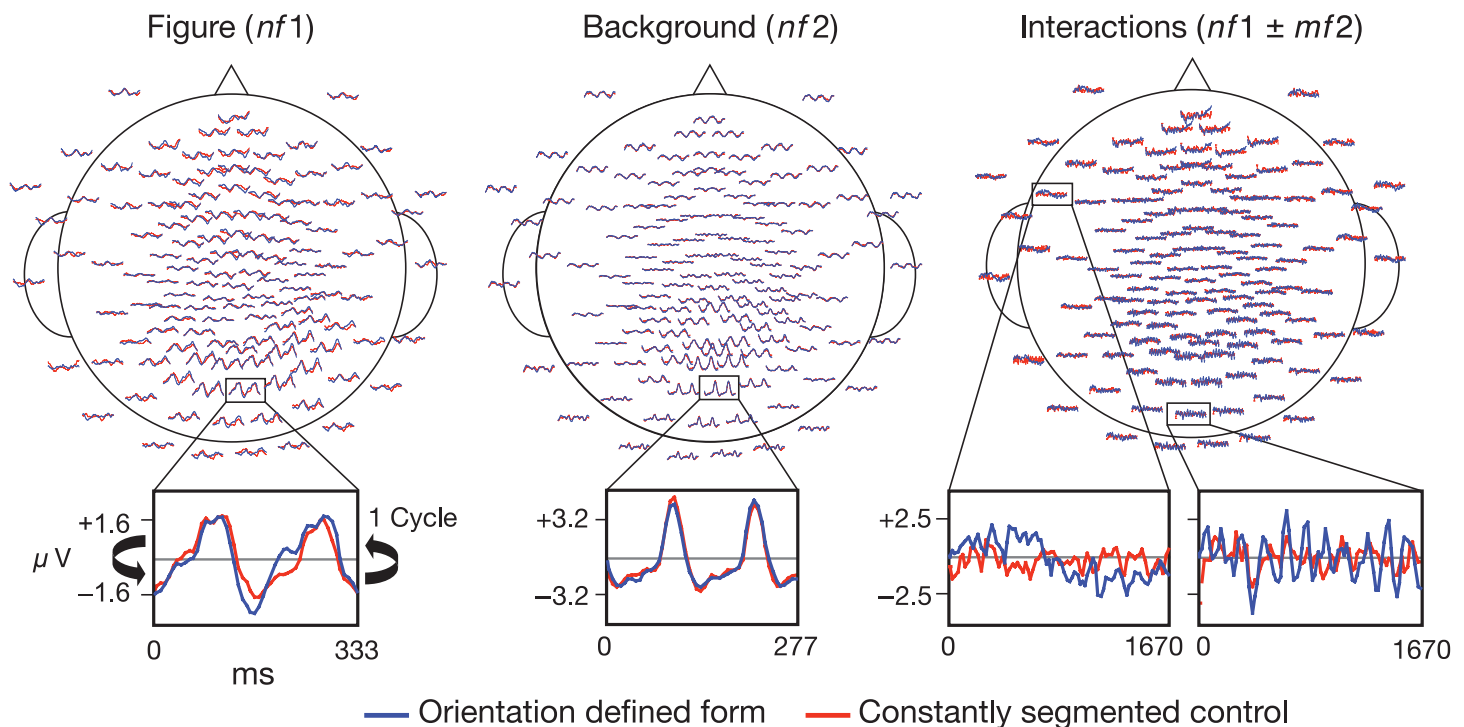


Figure 6. Re-analysis of orientation-defined form and control stimuli presented in Appelbaum et al., 2008 depicting grand average response waveforms, as in Figure 5 above. Waveform are shown for one cycle of the figure ($nf1$), the background ($mf2$), and the non-linear interactions ($nf1 \pm mf2$) for passively viewed orientation-defined form stimulus (blue) and texture segmentation control stimuli (red). The orientation-defined form stimuli are identical to those used in the present attentional variant and only differ in the subject's task. The control stimuli were identical in temporal structure but were composed of different random luminance textures in figure and background regions, and therefore the figure never blended into the background. Because a border was always present between regions, these stimuli are used to isolate aspects of processing that were specific to the appearance and disappearance of the segmented form. The observation that frontal and frontal-lateral, low-frequency bands were sensitive to border structure and task demands (see Figure 5) implies that these responses may be tapping common mechanisms involved in shape processing. See Supplementary Figures 5A, 5B, and 5C for enlarged images.

distributions for these and other response components are provided in Supplementary Figure 3.

Previous texture-segmentation VEP studies have reported an enhancement of late activity when the texture segmentation is task relevant (Bach & Meigen, 1992, 1997; Caputo & Casco, 1999; Casco et al., 2005; Fahle, Quenzer, Braun, & Spang, 2003; Heinrich et al., 2007; Khoe, Freeman, Woldorff, & Mangun, 2006; Schubo, Schroger, & Meinecke, 2004). In our displays, the dominant amplitude effects occur on the low-frequency interaction terms. There were no measurable amplitude effects of attention on the region responses themselves. It is thus possible the previously reported effects of attention on late components of the texture-segmentation VEP also arose from modifications of the non-linear interaction between stimulus regions. Conventional difference potential measures of texture-segmentation responses are unable to partition the response into separate region and region interaction components as we have done here and thus would not be able to make this distinction.

At present it is not clear what functional processes the two temporal bands of figure-ground interaction represent.

One possibility is a mechanism that extracts local discontinuities in feature maps occurring at object borders. We previously showed that the sum-term ($1f1 + 1f2$) is strongly dependent on the nature of the texture cues at the border of the two regions (Appelbaum et al., 2008). The sum-term and other high-frequency interaction terms are maximal in early visual areas. These areas are also known to contain end-stopped cells that would be expected to modulate strongly to transitions between continuous and discontinuous textures (Hubel & Wiesel, 1962). A re-analysis of data from Appelbaum et al. (2006) presented in Figure 6 shows that the low-frequency interaction components that occur outside of occipital areas also depend on the nature of the texture border information and also possibly on whether the global image changes between uniform and segmented states, or not. It should be noted, however, that non-linear spatial interactions that are not unique to texture segmentation paradigms (Fuchs, Andersen, Gruber, & Müller, 2008) have recently found intermodulation between rather widely separated flashing picture streams that were modifiable by task demands, under some conditions.

The sustained nature and stronger sensitivity to task demands of the low-frequency interactions in the present data suggests that they may be tapping more global mechanisms involved in shape processing, memory, recognition, or awareness. Previous MEG studies have suggested that low spatial frequency components of object photographs are routed to frontal areas and that these areas may send feedback to occipital object processing areas (Bar, 2003; Bar et al., 2006). This activity has been reported to involve very fast, beta-band oscillatory synchronization (Sehatpour et al., 2008), in contrast to the more sustained responses we have observed. This difference in dynamics notwithstanding, it is apparent that at least some aspects of object processing appear to involve frontal as well as occipital areas. Our results suggest one aspect of this frontal object network is involvement in task-dependent non-linear processing of both figure and background regions.

Integrating figure–ground segmentation and selective attention mechanisms

A critical step in the correct interpretation of visual scenes is the identification and assignment of borders between image regions that belong to different objects (Driver & Baylis, 1996; Nakayama & Mackeben, 1989). Single-unit studies have demonstrated that the representation of border ownership first occurs in cortical area V2 within 30–60 msec following the onset of a stimulus, with neurons coding the figure region's side of the border with an enhanced firing rate (Craft, Schutze, Niebur, & von der Heydt, 2007; Zhou, Friedman, & von der Heydt, 2000). Neurophysiological models of border ownership postulate that these processes are achieved via recurrent feedback from higher visual areas (Craft et al., 2007; Jehee, Roelfsema, Deco, Murre, & Lamme, 2007; Sakai & Nishimura, 2006; Zhaoping, 2005). Qui and colleagues (2007) have found that a subset of cells in V2 that code border ownership are also modulated by attention. In their study, the modulatory effect of attention was largest for cells coding the contour that was intrinsic to the attended figure. Activity in this population may be the basis of the larger task effects we observed on figure region as opposed to background region responses. It is not clear at this point whether the same population is involved in the effects we have observed on the figure–ground interaction. While cells in V2 are likely to contribute to the effects we observe at the occipital pole, they cannot account for the figure–ground activity we have observed at frontal sites. Focusing the task on the figure shape produced large modulations of frontal border related signals. These frontally distributed signals may contribute to the top-down control of figure region responses and to figure–ground interactions in earlier areas.

Lastly, as observed in previous variants of these designs that used similar figure–ground temporal tagging techniques

(Appelbaum et al., 2006, 2008), a basic spatial distinction of figure from background response distributions are maintained here that are independent of specific task demands. As seen before, the figure responses to these stimuli elicit a strongly bilateral occipital distribution, while the background produces a highly focal, midline spatial pattern. This fundamental processing distinction that occurs under highly demanding attentional allocation in both tasks is in agreement with numerous neurophysiological (Desimone & Duncan, 1995) and functional imaging (Kastner et al., 2000; Schira et al., 2004; Schubo et al., 2004) reports and illustrates that figure–ground organization precedes attentive processes. Consistent with this, some cells in V2 are border selective, independent of attention (Qiu, Sugihara, & von der Heydt, 2007).

Acknowledgments

Supported by INRSA EY14536 and EY06579 and the Pacific Vision Foundation. Special thanks to Marty Woldorff, William Wojtach, and Nico Boehler for their thoughtful comments on this manuscript.

Author contributions: Both authors contributed equally to this work.

Commercial relationships: none.

Corresponding author: Lawrence G. Appelbaum.

Email: greg@duke.edu.

Address: Center for Cognitive Neuroscience, Duke University, B203 LSRC, Box 90999, Durham, NC 27708, USA.

References

- Appelbaum, L. G., Wade, A. R., Pettet, M. W., Vildavski, V. Y., & Norcia, A. M. (2008). Figure–ground interactions in the human visual cortex. *Journal of Vision*, 8(9):8, 1–19, <http://journalofvision.org/8/9/8/>, doi:10.1167/8.9.8. [PubMed] [Article]
- Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W., & Norcia, A. M. (2006). Cue-invariant networks for figure and background processing in human visual cortex. *Journal of Neuroscience*, 26, 11695–11708. [PubMed] [Article]
- Bach, M., & Meigen, T. (1992). Electrophysiological correlates of texture segregation in the human visual evoked potential. *Vision Research*, 32, 417–424. [PubMed]
- Bach, M., & Meigen, T. (1997). Similar electrophysiological correlates of texture segregation induced by luminance, orientation, motion and stereo. *Vision Research*, 37, 1409–1414. [PubMed]

- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15, 600–609. [PubMed]
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 449–454. [PubMed] [Article]
- Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 8742–8747. [PubMed]
- Caputo, G., & Casco, C. (1999). A visual evoked potential correlate of global figure–ground segmentation. *Vision Research*, 39, 1597–1610. [PubMed]
- Casco, C., Grieco, A., Campana, G., Corvino, M. P., & Caputo, G. (2005). Attention modulates psychophysical and electrophysiological response to visual texture segmentation in humans. *Vision Research*, 45, 2384–2396. [PubMed]
- Chen, Y., Seth, A. K., Gally, J. A., & Edelman, G. M. (2003). The power of human brain magnetoencephalographic signals can be modulated up or down by changes in an attentive visual task. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3501–3506. [PubMed] [Article]
- Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure–ground organization. *Journal of Neurophysiology*, 97, 4310–4326. [PubMed] [Article]
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review Neuroscience*, 18, 193–222. [PubMed]
- Ding, J., Sperling, G., & Srinivasan, R. (2006). Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cerebral Cortex*, 16, 1016–1029. [PubMed] [Article]
- Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology*, 31, 248–306. [PubMed]
- Fahle, M., Quenzer, T., Braun, C., & Spang, K. (2003). Feature-specific electrophysiological correlates of texture segregation. *Vision Research*, 43, 7–19. [PubMed]
- Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision*, 8(1):12, 1–12, <http://journalofvision.org/8/1/12/>, doi:10.1167/8.1.12. [PubMed] [Article]
- Fuchs, S., Andersen, S. K., Gruber, T., & Müller, M. M. (2008). Attentional bias of competitive interactions in neuronal networks of early visual processing in the human brain. *NeuroImage*, 41, 1086–1101. [PubMed]
- Greenblatt, R. E., & Pflieger, M. E. (2004). Randomization-based hypothesis testing from event-related data. *Brain Topography*, 16, 225–232. [PubMed]
- Heinrich, S. P., Andres, M., & Bach, M. (2007). Attention and visual texture segregation. *Journal of Vision*, 7(6):6, 1–10, <http://journalofvision.org/7/6/6/>, doi:10.1167/7.6.6. [PubMed] [Article]
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154. [PubMed] [Article]
- Jehee, J. F., Roelfsema, P. R., Deco, G., Murre, J. M., & Lamme, V. A. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons-explaining crowding phenomena. *Brain Research*, 1157, 167–176. [PubMed]
- Kastner, S., De Weerd, P., & Ungerleider, L. G. (2000). Texture segregation in the human visual cortex: A functional MRI study. *Journal of Neurophysiology*, 83, 2453–2457. [PubMed] [Article]
- Khoe, W., Freeman, E., Woldorff, M. G., & Mangun, G. R. (2006). Interactions between attention and perceptual grouping in human visual cortex. *Brain Research*, 1078, 101–111. [PubMed]
- Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, 10, 117–125. [PubMed]
- Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, 20, 3310–3318. [PubMed]
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 8135–8139. [PubMed] [Article]
- Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 4770–4774. [PubMed] [Article]
- Müller, M. M., Andersen, S., Trujillo, N. J., Valdes-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 14250–14254. [PubMed] [Article]

- Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*, 631–634. [PubMed]
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*, 1631–1647. [PubMed]
- Pei, F., Pettet, M. W., & Norcia, A. M. (2002). Neural correlates of object-based attention. *Journal of Vision*, *2*(9):1, 588–596, <http://journalofvision.org/2/9/1/>, doi:10.1167/2.9.1. [PubMed] [Article]
- Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nature Neuroscience*, *10*, 1492–1499. [PubMed] [Article]
- Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.
- Sakai, K., & Nishimura, H. (2006). Surrounding suppression and facilitation in the determination of border ownership. *Journal of Cognitive Neuroscience*, *18*, 562–579. [PubMed]
- Schira, M. M., Fahle, M., Donner, T. H., Kraft, A., & Brandt, S. A. (2004). Differential contribution of early visual areas to the perceptual process of contour processing. *Journal of Neurophysiology*, *91*, 1716–1721. [PubMed] [Article]
- Scholte, H. S., Witteveen, S. C., Spekreijse, H., & Lamme, V. A. (2006). The influence of inattention on the neural correlates of scene segmentation. *Brain Research*, *1076*, 106–115. [PubMed]
- Schubo, A., Meinecke, C., & Schroger, E. (2001). Automaticity and attention: Investigating automatic processing in texture segmentation with event-related brain potentials. *Brain Research: Cognitive Brain Research*, *11*, 341–361. [PubMed]
- Schubo, A., Schroger, E., & Meinecke, C. (2004). Texture segmentation and visual search for pop-out targets. An ERP study. *Brain Research: Cognitive Brain Research*, *21*, 317–334. [PubMed]
- Sehatpour, P., Molholm, S., Schwartz, T. H., Mahoney, J. R., Mehta, A. D., Javitt, D. C., et al. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 4399–4404. [PubMed] [Article]
- Victor, J. D., & Mast, J. (1991). A new statistic for steady-state evoked potentials. *Electroencephalography Clinical Neurophysiology*, *78*, 378–388. [PubMed]
- Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area v2. *Neuron*, *47*, 143–153. [PubMed]
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*, 6594–6611. [PubMed] [Article]